DOI:10.19651/j.cnki.emt.2208892

基于平行图像的糖尿病视网膜病变智能诊断*

赵 亮[1] 付园坤[1] 陈涵欣[1] 魏政杰[1] 云 晴[1] 金军委[2]
(1.河南工业大学电气工程学院 郑州 450001;2.河南工业大学大数据与人工智能学院 郑州 450001)

摘 要: 针对深度学习诊断糖尿病视网膜病变(DR)面临数据集小、类别不均衡及诊断效果不佳等问题,提出基于平行图像和 Swin Transformer 的 DR 分级模型。首先基于 StylcGAN2-ada 构建平行图像生成模型,解决训练图像过少和类别失衡问题。经 FID、KID 和目视评估,构建的平行图像符合后续工作要求。然后,基于注意力与窗口滑动机制构建 DR 诊断模型改善诊断效果。最后,使用平行图像训练诊断模型。经验证,本文提出的诊断模型准确率为 93.5%、特异性最高为 99%、F1 分数最高为 0.96。与原始图像相比,使用平行图像训练模型后其准确率提升 20%、精确率最高提升 70%。与其他 3 种深度学习模型对比,本文所提方法各项指标均达到最优。以上结果表明,本文构建的模型可在小样本数据集下实现较好的诊断效果。

关键词: 糖尿病视网膜病变;平行眼底图像;生成对抗网络;Swin Transformer
中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Intelligent diagnosis of diabetic retinopathy based on parallel images

Zhao Liang[1] Fu Yuankun[1] Chen Hanxin[1] Wei Zhengjie[1] Yun Qing[1] Jin Junwei[2]
(1. College of Electrical Engineering, Henan University of Technology, Zhengzhou 450001, China;
2. School of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou 450001, China)

Abstract: Aiming at the problems of small datasets, unbalanced categories and poor diagnostic results in deep learning diagnosis of diabetic retinopathy (DR), this paper proposes a DR grading model based on parallel images and Swin Transformer. First, build a parallel image generation model based on StyleGAN2-ada to solve the problem of too few training images and class imbalance. After FID, KID and visual evaluation, the constructed parallel images meet the requirements of subsequent work. Then, a DR diagnosis model is constructed based on the attention and window shifting mechanism to improve the diagnosis effect. Finally, a diagnostic model is trained using the parallel images. After verification, the accuracy of the diagnostic model proposed in this paper is 93.5%, the highest specificity is 99%, and the highest F1-score is 0.96. Compared with the original images, the accuracy of the model is improved by 20% and the accuracy is improved by up to 70% after training the model with parallel images. Compared with the other three deep learning models, all the indicators of the method proposed in this paper are optimal. The above results show that the model constructed in this paper can achieve better diagnostic results under a small sample data set.

Keywords: diabetic retinopathy; parallel fundus image; generative adversarial network; Swin Transformer

0 引 言

糖尿病视网膜病变(diabetic retinopathy, DR)是一种常见的糖尿病并发症,会对患者视力造成严重危害。根据国际糖尿病联盟调查,2021 年全球约有 5.37 亿糖尿病患者,其中 1/3 患有 DR[1]。目前眼科医生主要借助眼底视网膜照相、眼底血管造影、光学相干断层扫描和超声检查等辅助诊断 DR[2]。人工方法虽已取得一定成功,但是诊断结果主观性较强,另外,培养一名合格的眼科医生需要大量成本,造成经济欠发达地区患者无法及时治疗。针对以上人工诊断存在的问题,业界尝试研发一种低成本便捷高效的诊断方案。

随着计算机技术的进步,一些学者开始尝试使用计算机和自动化技术辅助诊断 DR。视网膜图像中血管提取对眼科疾病诊断至关重要,苑玮琦等[3]为改善低对比度视网膜血管提取效果差的情况,提出了基于主曲率和主方向的

收稿日期:2022-01-19
* 基金项目:国家自然科学基金(61473114,62106068)项目资助

血管骨架提取法，该方法能够提取出微小血管和低对比度血管。不同于以上方法，Lachure 等[4]提出通过使用支持向量机和 K 最邻近算法检测微动脉瘤、出血和软硬渗出物，从而实现 DR 自动筛查，实验表明支持向量机比 K 最邻近分类有更好的效果。为了进一步提高分级精确率，Reddy 等[5]使用决策树和逻辑回归等建立集成学习模型诊断 DR，实验表明该模型的精确率、召回率和 F1 分数等指标优于单个机器学习算法。以上技术需要使用手工设计得到视网膜图像特征效率较低，此外在不同采集条件下获得的视网膜图像差异较大，影响模型的诊断效果。

近年来芯片技术的进步为深度学习提供了强大的算力，基于深度学习的 DR 诊断开始成为研究热点。Saranya 等[6]提出使用卷积神经网络构建 DR 分级系统，该模型可以将 DR 病变分为 4 个等级，在 Mcssidor 数据集上的准确率、特异性和敏感性均达到了 88%以上。区别于以上模型的输出结果，Amalia 等[7]提出使用卷积神经网络与长短期记忆网络结合的方法检测 DR，首先使用 GoogLeNet 提取图像特征，然后把图像信息和描述语句输入长短期记忆网络中，最后输出包含诊断结果的语句，其准确率达到了 89.65% 以上。张思杰等[8]提出使用生成对抗网络(generative adversarial networks，GAN)的少样本视网膜血管分割方法，生成器使用 U-net 结构，判别器使用卷积神经网络，最终的曲线下面积达到 0.95 以上，准确率高于 94%。总结现有成果可以看到，基于深度学习的 DR 诊断需要大量高质量视网膜图像进行模型训练，而医学图像因涉及个人隐私问题获取困难，且图像标注需要专业的眼科医生完成成本较高。此外眼底图像的解剖结构复杂细节丰富，传统的卷积操作不能高效聚焦病灶特征，影响诊断结果的准确率和鲁棒性。

针对现有问题，本文提出使用少量视网膜图像的 DR 分级方法。首先对视网膜图像数据集进行处理，然后基于生成对抗网络 StyleGAN2-ada[9]构造生成模型生成虚拟眼底图像，它和原始真实图像混合得到平行图像。接着使用平行图像构建基于 Swin Transformer[10]的智能诊断模型，该模型的窗口滑动机制和注意力机制能高效提取图像细节信息，节省计算资源的同时提高模型的诊断准确率。最后通过 3 组实验验证本文提出方法，即：基于 StyleGAN2-ada 模型的平行图像生成实验、基于平行图像的 Swin Transformer 诊断模型实验和对比实验。结果表明，本文构建的模型有效解决了数据量较少的问题，提高了 DR 诊断准确率，为 DR 智能诊断提供新方法。

本文的主要贡献有两点：1)针对 DR 视网膜图像获取困难和类别不平衡问题提出使用 StyleGAN2-ada 生成虚拟图像，然后与原始图像组成平行图像提高模型诊断效果；2)构建基于注意力机制和滑动窗口机制的 Swin Transformer DR 诊断模型，该模型可有效地关注并提取 DR 图像中的病灶细节信息，节省计算资源，提升诊断性能。

1　平行视网膜图像生成模型构建

1.1　模型构建

用于训练 DR 智能诊断模型的视网膜图像涉及患者隐私难以大量获取，而深度学习模型的性能又极度依赖大量高质量的训练样本。为克服这一矛盾，本文提出使用生成对抗网络 StyleGAN2-ada 生成与真实视网膜图像概率分布一致的虚拟图像，然后与真实图像一起作为 DR 诊断模型的训练样本。它根据输入的潜在编码和随机噪声输出虚拟视网膜图像，判别器鉴别输入图像的真伪。在训练过程中生成器和判别器相互博弈，对抗中优化各自模型直至纳什均衡。它的目标函数可以概括为以下形式：

$$\min_G \min_D V(G,D) = \min_G \min_D V(G,D) E_{x\sim p_{data}}[\log D(x)] + E_{z\sim p_z}[\log(1-D(G(z))] \tag{1}$$

式中：z 表示随机噪声，x 为真实图像，G 和 D 分别表示生成器和判别器。该函数表示令生成样本概率分布和真实样本概率分布之间的 KL 散度达到最小。

传统的生成对抗网络生成过程是在“黑箱”中进行[11]，无法理解随机噪声如何控制图像生成，另外 GAN 在训练时为防止模型过拟合通常会对图像进行翻转、加入噪声和色彩强化等图像增强操作，使得传统 GAN 生成的图像也存在增强效果。相较之下，StyleGAN2-ada 实现了图像风格可控，并且能使用少量真实样本生成高质量虚拟图像。它由生成器和判别器两个部分组成，生成器结构与 StyleGAN 类似(如图 1 所示)共有 35 层，包括输入输出层、映射网络和生成模块。首先，生成器输入潜在编码，通过由 8 层全连接层组成的非线性网络映射层映射至 W 空间，然后再输入至各个生成器。虚拟视网膜图像由 7 个生成模块生成，第 1 个 4×4 的生成模块相比其它生成模块多一层输入层，这些常数和映射网络的输出 w 与噪声 B 组成生成模块的输入。图像生成的本质是上采样过程，StyleGAN 使用自适应实例规范化(adaptive instance normalization，AdaIN)控制生成图像的风格，然后在卷积操作之后输入噪声 B，生成模块中 AdaIN 操作表达式如下：

$$\mathrm{AdaIN}(x_i,y) = y_{s,i}\frac{x_i-\mu(x_i)}{\sigma(x_i)} + y_{b,i} \tag{2}$$

其中，$y=(y_s,y_b)$ 表示非线性映射的 DR 图像的细节特点，x_i 是第 i 层的特征图，$\mu(x_i)$ 表示各通道的特征图全部像素的均值，$\sigma(x_i)$ 表示各通道的特征图全部像素的方差。经过第一个生成模块生成 4×4 分辨率的图像后，图像以两倍的趋势增长依次上采样至 8×8、16×16 等分辨率，6 次上采样后生成 256×256 分辨率的视网膜图像，最后通过一个卷积层输出 RGB 图像。

StyleGAN2-ada 的特点是输入的潜在变量非线性映射至 W 空间后再输入至各生成模块，该结构能够丰富生成的视网膜图像细节如血管、视盘、渗出液等。另一方面，判别器使用了自适应鉴别增强机制(如图 2 所示)避免生成的虚

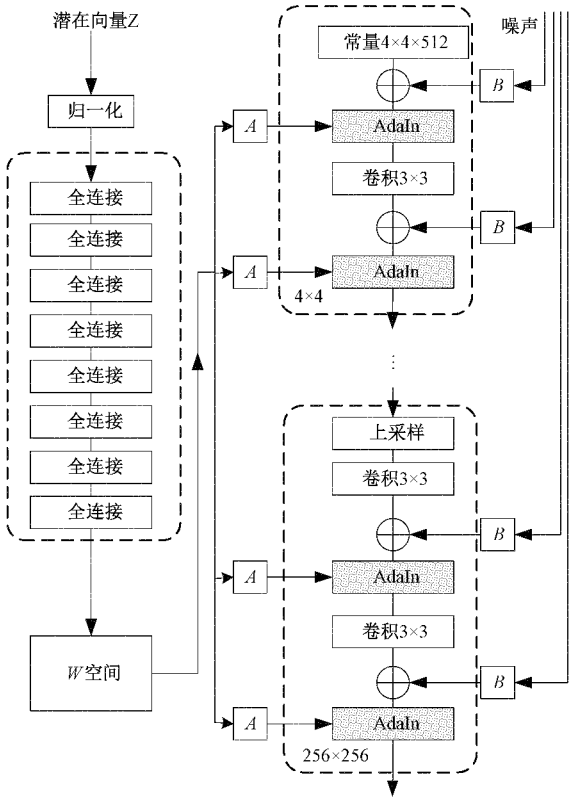

图 1　StyleGAN 生成器结构示意图

拟图像存在增强化现象,其中 $f(x)=\log(sigmoid(x))$,通过网络自适应机制每 4 个批次调整一次图像增强的概率 p,此概率决定是否对图像进行增强操作。判别器的本质是将图像下采样后通过一个全连接层进行真实/虚拟分类输出相应概率。本文中判别器共 25 层将输入图像从 256×256 分辨率通过 6 次卷积降低采样率至 4×4,每次分辨率降低至原来的 1/2,通过全连接层输出 DR 视网膜图像的判别结果。模型训练完成后,先生成虚拟图像再与真实图像混合得到平行图像。

1.2　生成的虚拟视网膜图像质量评价方法

本文使用 FID(fréchet inception distance)和 KID(kernel inception distance)[12-13]以及目视 3 个指标评估生成图像的质量,其中 FID 和 KID 是在 IS(inception score)的基础上进行改进。IS 使用 InceptionNet-V3[14]对生成图像进行分类,如果它能够较高概率的预测图像中包含的真实物体,说明图像质量达到一定高度,该指标表达如下:

$$IS(G)=\exp(E_{x\sim p_g}D_{\mathrm{KL}}(p(y\mid x)\parallel p(y))) \tag{3}$$

$$p(y)=\frac{1}{N}\sum_{i=1}^{N}p(y\mid x^{(i)}) \tag{4}$$

式中:p_g 是生成的虚拟视网膜图像的概率分布,$p(y\mid x)$ 是生成图像 x 属于各类别的概率,$p(y)$ 是在所有类别上的边缘分布,$D_{\mathrm{KL}}(\cdot\parallel\cdot)$ 是 $p(y\mid x)$ 和 $p(y)$ 的 KL 散度。

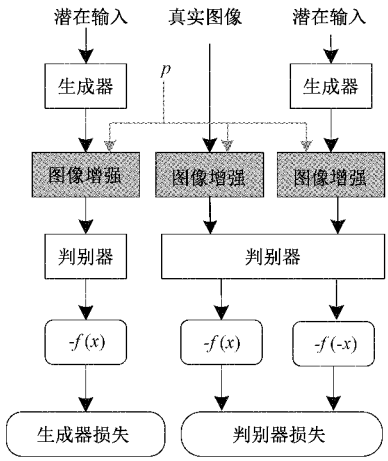

图 2　判别器结构示意图

因为 IS 指标无法反映真实图像和生成图像之间的距离,FID 提出了改进方法。具体地,FID 使用 Inception 网络提取图像特征然后使用高斯模型对特征空间建模计算两个特征之间 Wasserstein-2 距离。

$$FID(x,g)=\parallel\mu_x-\mu_g\parallel_2^2+Tr(\sum_x+\sum_g-2(\sum_x\sum_g)^{\frac{1}{2}}) \tag{5}$$

其中,x 是真实图像,g 是生成图像,μ 代表对应图片的特征均值,$Tr(\cdot)$ 是矩阵的迹,$\sum$ 表示方差。在评价生成图像质量时,FID 越小表示生成的视网膜图像越好。KID 指数与 FID 类似,通过计算 Inception 网络表征之间的最大均值差异度量图像概率分布之间的差异,KID 越小表示生成图像质量越好。

2　基于平行图像的诊断模型构建

2.1　模型构建

DR 视网膜图像细节丰富,目前使用深度卷积网络的诊断方法不能高效地关注图像细节信息,有些深度学习诊断模型甚至存在随着网络层数增多图像细节信息丢失的情况。针对此问题,本文使用基于注意力机制的 Swin Transformer 构建 DR 诊断模型。它是一种改进型的 Transformer[15]架构的计算机视觉模型,在 COCO 数据集[16]上进行目标检测和实例分割都取得了优异成绩。与自然语言处理任务不同,Transformer 处理图像任务时面临信息量较大的问题。Swin Transformer 为解决这一问题提出分层的 Transformer 结构,将自注意力操作限制到不重叠的窗口内,并使用平移窗口方法连接上一层窗口(如图 3 所示)。这种机制增强窗之间特征图的连接,有利于重要特征的提取。

本文构建的 DR 智能诊断模型如图 4 所示,平行 DR 图像首先经过卷积核与步长都为 4 的卷积操作被分为若干个分辨率为 56×56 的图像块,对图像块归一化和 dropout 处理后进行特征提取,特征提取为 4 个阶段组成的网络。

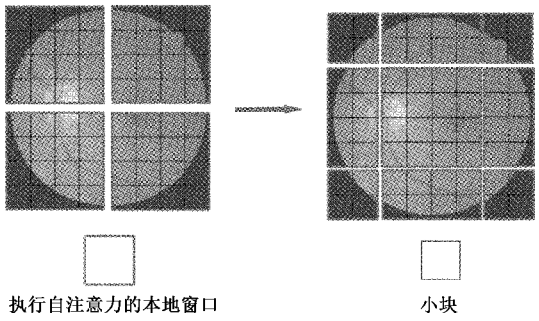

图 3　Swin Transformer 的滑动窗口注意力

其中第 1、2、4 阶段的深度均是 2，第 3 阶段的深度是 6，且 1～4 阶段的注意力头数分别为 3、6、12、24。从图 4 可知，2～4 阶段输入 Swin Transformer 块前进行合并操作，每经过一个阶段图像分辨率降低至原来的 1/2，第 4 阶段图像分辨率降低至 7×7，最后经过平均池化和线性分类输出病变等级。每个阶段中 Swin Transformer 块的结构如图 5 所示，其中 LN 表示层归一化，MLP 是多层感知器，相邻块之间分别使用了基于窗口的多头自注意力(window based multi-head self-attention，W-MSA)和窗口滑动的多头自注意力(shifted window-based multi-head self-attention，SW-MSA)模块，Swin Transformer 块的计算可以表示为：

$$\begin{aligned}\hat{z}l &= \mathrm{W-MSA}(\mathrm{LN}(z^{l-1})) + z^{l-1}\\ z^{l} &= \mathrm{MLP}(\mathrm{LN}(\hat{z}l)) + \hat{z}l\\ \hat{z}^{l+1} &= \mathrm{SW-MSA}(\mathrm{LN}(z^{l})) + z^{l}\\ z^{l+1} &= \mathrm{MLP}(\mathrm{LN}(\hat{z}^{l+1})) + \hat{z}^{l+1}\end{aligned} \tag{6}$$

其中，z^{l} 和 $\hat{z}l$ 分别是第 l 块 MLP 和(S)W-MSA 的输出，W-MSA 和 SW-MSA 中自注意力的计算与 Transformer 一致。

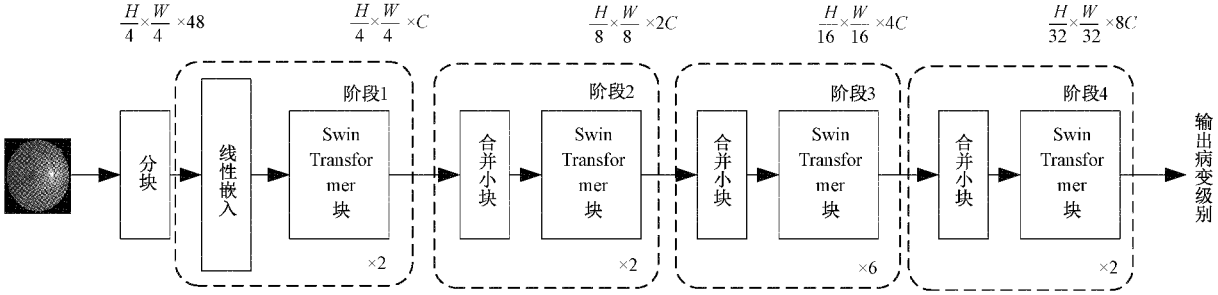

图 4　DR 智能诊断模型

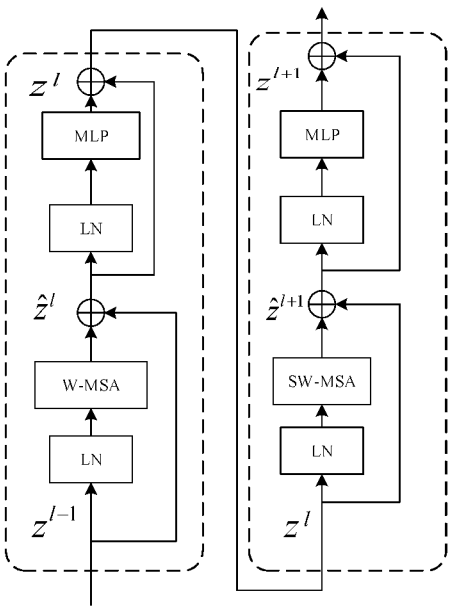

图 5　Swin Transformer 块

2.2　诊断模型性能评价方法

由 DR 图像划分病变等级的本质是图像分类问题，根据样本的类别与分类结果有 4 种情况：真阳性(true positive，TP)、真阴性(true negative，TN)、假阳性(false positive，FP)和假阴性(false negative，FN)。其中真阳性表示诊断结果和样本均为阳性；真阴性表示诊断结果和样本均为阴性；假阳性表示诊断结果为阳性，但样本为阴性；假阴性表示诊断结果为阴性，但样本为阳性。

对于 DR 分级模型，通常用精确率(precision)、准确率(accuracy)、敏感性(sensitivity)、特异性(specificity)、各类别 F1 分数(如表 1 所示)和混淆矩阵评估模型的性能。混淆矩阵热力图能够可视化评价 DR 诊断模型的分级情况，在混淆矩阵中主对角线方向的色块颜色越深，说明模型性能越好。

表 1　DR 智能诊断模型评价指标

名称	表达式
精确率	$\dfrac{TP}{TP+FP}$
准确率	$\dfrac{TP+TN}{TP+FP+TN+FN}$
敏感性	$\dfrac{TP}{TP+FN}$
特异性	$\dfrac{TN}{TN+FP}$
各类别 F1 分数	$2\cdot\dfrac{sensitivity\cdot precision}{sensitivity+precision}$

3　实验设置

本文设置了 3 组实验验证提出的 DR 诊断方法:实验 1,验证基于 StyleGAN2-ada 模型的生成效果;实验 2,验证基于平行图像的 Swin Transformer 模型诊断效果;实验 3,对比使用平行图像与原始图像分别构建的 Swin Transformer 诊断模型效果,然后比较使用平行图像时 Swin Transformer 与其他深度学习模型的诊断效果。

在进行实验验证时使用的数据集、实验环境和参数设置如下。

3.1　实验数据集

本实验使用公开的 Messidor 数据集[17]生成虚拟图像,该数据集由 3 个不同的眼科部门采集,图像格式为 tif,包括 1 440×960、2 240×1 488 和 2 304×1 536 三种分辨率,每张眼底彩色图像有黄斑水肿风险等级和糖尿病视网膜病变等级两种标签。它共有各类型视网膜图像 1 200 张,除去错误标注和重复图像,有效样本图像 1 187 张,各个等级图像数量的占比如表 2 所示。

表 2　视网膜图像数据集

等级	数量	比例/%
正常(0)	547	46
轻度(1)	149	13
中度(2)	240	20
重度(3)	251	21

该数据集根据眼底图像的临床表现如微动脉瘤、出血和新生血管的位置和数量划分为正常、轻度、中度和重度 4 个等级[18]。

3.2　实验环境

本文在 Tensorflow1.14 框架下实现虚拟视网膜图像生成,运行程序的服务器搭载了 8 块 Nvidia Geforce 2080Ti GPU、1 块 Intel Xeon Silver 4214 CPU 和 512 G 内存。另外,使用 Pytorch3.7 框架实现基于平行图像的 Swin Transformer 诊断模型,运行程序的台式机使用 1 块 Nvidia Geforce 2080Ti GPU、1 块 Intel i7 9700K CPU 和 32 G 内存。

3.3　参数设置

本文使用 4 种等级视网膜图像分别训练 StyleGAN2-ada 模型,初始学习率为 2×10^{-4},当模型的 FID 值降低至 25 时,逐步调整学习率进一步降低 FID 值,直至 FID 趋于稳定则训练结束。然后使用混合后的 4 494 张平行图像训练 Swin Transformer 诊断模型,各等级图像数目如表 3 所示,训练时使用学习率递减策略,初始学习率是 1×10^{-4},模型优化器使用 AdamW(adam with decoupled weight decay),批次大小设定为 8,最大迭代 300 次。观察模型在训练集和验证集上的性能表现,如果模型的损失和准确率在 300 次迭代结束时收敛,则训练完成。在进行对比实验时,训练 MobileNetV2 使用自适应矩估计(adaptive moment estimation, Adam)优化器,而 EfficientNetV2 和 Vision Transformer 使用随机梯度下降法进行参数更新,对比实验的最大迭代次数均为 300。

4　实验结果与分析

4.1　平行视网膜图像生成结果与分析

在平行图像生成阶段,将 Messidor 数据集进行如下处理:1)剔除重复图像,修正错误标签;2)将图像分辨率下调至 256×256,加快 StyleGAN2-ada 模型的训练速度;3)将图像格式转化成 png;4)对图像进行翻转、平移、颜色增强等操作;5)使用 StyleGAN2-ada 生成虚拟视网膜图像;6)混合虚拟图像与真实图像得到平行图像。

本文 FID 度量 5 000 张真实图像和虚拟图像间的距离,为了方便观察,将 KID 放大 1 000 倍记为 $KID_{\times1\,000}$。生成模型训练 0 等级图像时 FID 值的变化趋势如图 6 所示,可以看出随着训练的进行生成图像的 FID 值逐渐收敛稳定在 17 左右,表明模型构建得当且训练充分。本文基于指标和目视分析生成模型的性能,各等级生成虚拟视网膜图像的 FID 和 $KID_{\times1\,000}$ 数值如表 3 所示,结果表明各等级图像指标均趋于一致,其中正常图像的 KID 值偏大,分析原因可能是原始图像中正常等级的图像数量最多,在进行特征距离度量时会根据图像数量相应增大;相对地,轻度等级的图像最少,因此 KID 值最小。通过目视观察对比图 7 中的虚拟图像与真实图像,可以发现模型生成的虚拟图像轮廓清晰,视盘、视杯、静脉和动脉边界清楚,黄斑可见,亮度和真实视网膜图像高度一致。

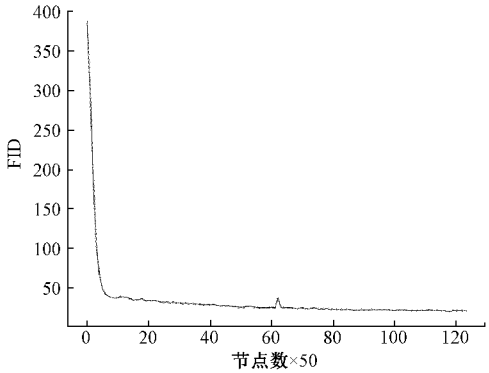

图 6　训练 0 等级图像生成模型时 FID 的变化

表 3　基于 StyleGAN2-ada 的生成模型性能

等级	原始数据集图像数量	平行图像数量	*FID*	$KID_{\times1\,000}$
正常	547	1 496	17.1	12.150
轻度	149	1 095	17.8	2.842
中度	240	1 176	16.7	4.240
重度	251	1 177	19.4	5.663

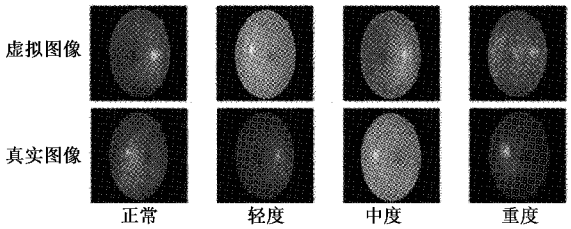

图 7　虚拟与真实视网膜图像

综上所述，构建的 StyleGAN2-ada 生成模型较好的学习了真实视网膜图像数据的概率分布，生成的虚拟图像质量符合后续工作要求。

4.2　基于平行图像的诊断模型结果与分析

本文构建的基于平行图像的糖尿病视网膜病变诊断模型表现出色。具体地，使用 StyleGAN2-ada 生成的虚拟视网膜图像，混合原始图像后得到 4 944 张平行视网膜图像。然后使用平行图像对基于 Swin Transformer 构建的诊断模型进行训练和验证，训练过程如图 8 所示，迭代优化 60 次时，Swin Transformer 的诊断准确率达到 91.9%。最后经过 300 次迭代训练，模型损失收敛至 0.1 附近，准确率最高达到 93.52%。诊断模型的精确率、敏感性、特异性和 F1 分数如表 4 所示，不同等级病变的诊断精确率、敏感性和特异性最高均达到了 96%，F1 分数最高达到 0.96。

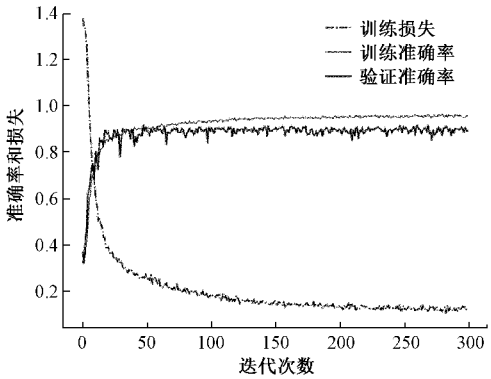

图 8　基于平行图像的诊断模型训练过程

表 4　基于平行图像的诊断模型性能

等级	精确率	敏感性	特异性	各类别 F1 分数
正常	0.903	0.963	0.955	0.932 0
轻度	0.961	0.909	0.990	0.934 2
中度	0.934	0.898	0.980	0.915 6
重度	0.958	0.962	0.987	0.959 9

混淆矩阵热力图（图 9）右侧的色带表示被准确分类的程度，颜色越深表示被正确分类的样本越多，由混淆矩阵可以看出各等级图像被准确分类的效果较好。另外，正常和重度等级病变的诊断效果好于轻度和中度的效果，原因可能是正常和重度等级的图像相对较多，模型训练较充分，也印证了数据量对模型结果影响较大。综上可以看出基于平行图像的诊断模型表现出色符合预期。

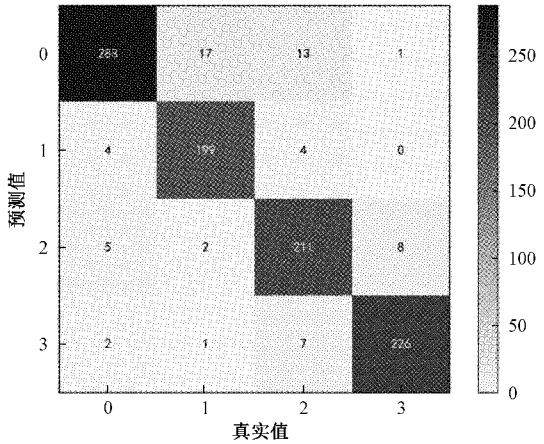

图 9　基于平行图像的诊断模型混淆矩阵

4.3　对比实验结果与分析

这部分通过两组对比实验结果进一步验证本文所提方法的性能。

1）实验 1：对比使用平行图像与原始图像分别构建的 Swin Transformer 诊断模型效果

本实验目的是比较使用平行图像和原始图像分别构建 Swin Transformer 诊断模型性能的差异。原始数据集由 1 187 张图像组成，其中 80%作为训练样本，剩余的作为验证样本。基于原始图像的 Swin Transformer 诊断模型实验结果如表 5 所示，该模型准确率达到 70.76%，精确率最高为 81.6%，轻度等级病变诊断精确率只达到 27.3%，与表 4 相比，使用原始图像的诊断模型性能严重退化。分析图 10 的混淆矩阵热力图可以看出图中各色块分布散乱，轻度和中度病变诊断效果较差。对比 4.2 节的实验结果表明平行图像增加了样本的数量与多样性，消除了类别不均衡的问题，明显改善诊断模型效果。

表 5　基于原始图像 Swin Transformer 的诊断模型结果

等级	精确率	敏感性	特异性	各类别 F1 分数
正常	0.759	0.927	0.748	0.834
轻度	0.273	0.103	0.961	0.149
中度	0.535	0.479	0.894	0.505
重度	0.816	0.800	0.952	0.807

2）实验 2：对比使用平行图像时 Swin Transformer 与其他深度学习模型的诊断效果

本实验目的是对比使用平行图像时 Swin Transformer 与 MobileNetV2[19]、EfficientNetV2[20] 以及 Vision Transformer[21] 诊断模型性能的差异。对比试验各模型均训练 300 轮，准确率如图 11 所示，可以看出本文基于平行图像和 Swin Transformer 构建的诊断模型收敛速度较快，准确率优于其他模型。各等级病变的诊断精确率、敏感性

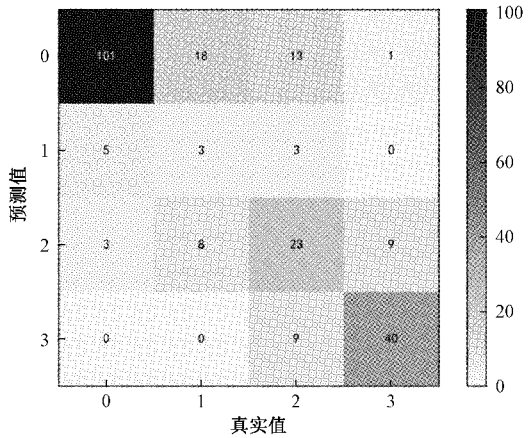

图 10　基于原始图像 Swin Transformer 诊断模型的混淆矩阵

和特异性对比如表 6、7 和 8 所示,其余 3 种模型表现均不如本文提出的模型,另外混淆矩阵(图 12、13 和 14)也验证了以上结论。

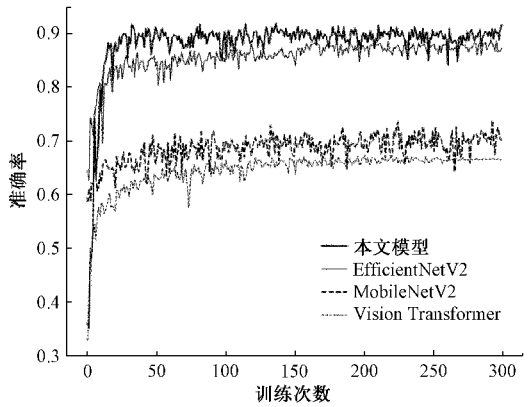

图 11　本文诊断模型与其他深度学习模型的准确率对比

表 6　精确率对比

等级	Swin Transformer	Vison Transformer	Efficient NetV2	Mobile NetV2
正常	0.903	0.546	0.316	0.859
轻度	0.961	0.543	0.229	0.682
中度	0.934	0.322	0.252	0.600
重度	0.958	0.511	0.236	0.850

表 7　敏感性对比

等级	Swin Transformer	Vison Transformer	Efficient NetV2	Mobile NetV2
正常	0.963	0.649	0.251	0.836
轻度	0.909	0.174	0.333	0.685
中度	0.898	0.413	0.268	0.740
重度	0.962	0.570	0.183	0.677

表 8　特异性对比

等级	Swin Transformer	Vison Transformer	Efficient NetV2	Mobile NetV2
正常	0.955	0.766	0.765	0.940
轻度	0.990	0.958	0.680	0.909
中度	0.980	0.729	0.752	0.846
重度	0.987	0.830	0.815	0.963

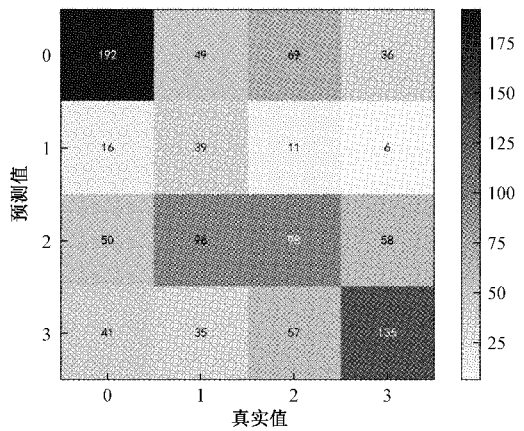

图 12　基于平行图像和 VisionTransformer 模型的混淆矩阵

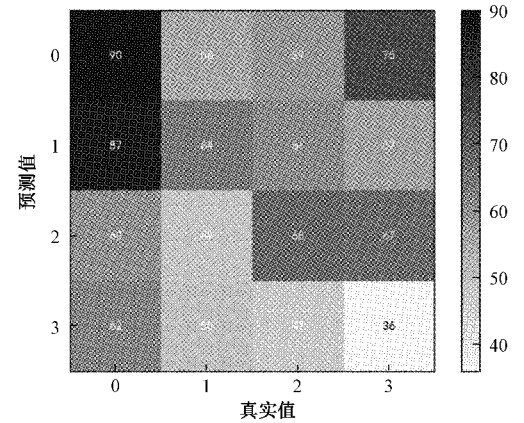

图 13　基于平行图像和 EfficientNetV2 诊断模型的混淆矩阵

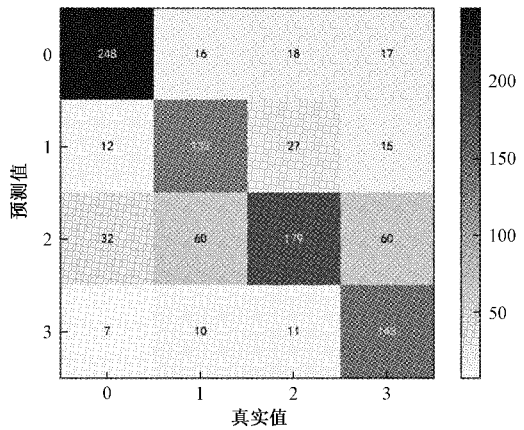

图 14　基于平行图像和 MobileNetV2 模型的混淆矩阵

本文提出的基于平行图像和 Swin Transformer 构建的诊断模型在精确率、敏感性和特异性等方面均高于其他深度学习模型，原因是本模型的注意力和窗口滑动机制可有效提取局部病灶信息并构造病变视网膜图像的全局特征，从而提高模型的诊断性能。

5　结　　论

本文提出一种基于平行图像的 DR 智能诊断方法，根据少量的原始真实视网膜图像，使用基于 StyleGAN2-ada 的生成模型产生与真实图像分布一致的虚拟图像，进而得到平行图像，然后使用该平行图像构建 Swin Transformer 诊断模型实现 DR 的严重程度分级。结果表明本模型精确率、敏感性和特异性均达到 96%以上，F1 分数最高达到 0.96。该方法能解决训练图像过少和类别不均衡问题，改善数据量较少导致 DR 智能诊断准确性无法保证的情况。注意力和滑动窗口机制的引入能高效提取视网膜图像中的细节特征，从而提高模型的收敛速度和诊断性能。

为进一步提高诊断模型的准确率和鲁棒性，可考虑集成 Vision Transformer、Swin Transformer 和残差网络等多个异质深度学习模型作为诊断工具进行 DR 筛查转诊。另一方面，也可推广该方法至数据样本量较少的罕见病智能诊断任务。

参考文献

[1] SUN H, SAEEDI P, KARURANGA S, et al. IDF diabetes atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045[J]. Diabetes Research and Clinical Practice, 2022, 183: 109119.

[2] 高斐，闵寒毅. 糖尿病视网膜病变的诊断与治疗[J]. 中国临床医生杂志，2021，49(12)：1402-1404.

[3] 苑玮琦，王安. 基于主曲率和主方向的多尺度视网膜血管骨架提取方法[J]. 仪器仪表学报，2021，42(6)：191-199.

[4] LACHURE J, DEORANKAR A V, LACHURE S, et al. Diabetic retinopathy using morphological operations and machine learning [C]. 2015 IEEE International Advance Computing Conference(IACC), IEEE, 2015: 617-622.

[5] REDDY G T, BHATTACHARYA S, RAMAKRISHNAN S S, et al. An ensemble based machine learning model for diabetic retinopathy classification[C]. 2020 International Conference on Emerging Trends in Information Technology and Engineering(ic-ETITE), IEEE, 2020: 1-6.

[6] SARANYA P, PRABAKARAN S. Automatic detection of non-proliferative diabetic retinopathy in retinal fundus images using convolution neural network[J]. Journal of Ambient Intelligence and Humanized Computing, 2020: 1-10.

[7] AMALIA R, BUSTAMAM A, SARWINDA D. Detection and description generation of diabetic retinopathy using convolutional neural network and long short-term memory[C]. Journal of Physics: Conference Series, IOP Publishing, 2021, 1722(1): 012010.

[8] 张思杰，方翔，魏赋. 基于 GAN 的少样本视网膜血管分割研究[J]. 电子测量与仪器学报，2021，35(11)：132-142.

[9] KARRAS T, AITTALA M, HELLSTEN J, et al. Training generative adversarial networks with limited data[J]. ArXiv Preprint, 2020, ArXiv: 2006.06676.

[10] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[J]. ArXiv Preprint, 2021, ArXiv: 2103.14030.

[11] 王延年，李文婷，任劼. 基于生成对抗网络的单帧图像超分辨算法[J]. 国外电子测量技术，2020，39(1)：26-32.

[12] OBUKHOV A, KRASNYANSKIY M. Quality assessment method for GAN based on modified metrics inception score and fréchet inception distance [C]. Proceedings of the Computational Methods in Systems and Software, Springer, Cham, 2020: 102-114.

[13] BINKOWSKI M, SUTHERLAND D J, ARBEL M, et al. Demystifying mmd gans[J]. ArXiv Preprint, 2018, ArXiv: 1801.01401.

[14] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2818-2826.

[15] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]. Advances in Neural Information Processing Systems, 2017: 5998-6008.

[16] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context[C]. European Conference on Computer Vision. Springer, Cham, 2014: 740-755.

[17] DECENCIÈRE E, ZHANG X, CAZUGUEL G, et al. Feedback on a publicly distributed image database: The messidor database[J]. Image Analysis & Stereology, 2014, 33(3): 231-234.

[18] BHARDWAJ C, JAIN S, SOOD M. Diabetic retinopathy severity grading employing quadrant-based Inception-V3 convolution neural network architecture[J]. International Journal of Imaging Systems and Technology, 2021, 31(2): 592-608.

[19] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.

[20] TAN M, LE Q V. Efficientnetv2: Smaller models and faster training [J]. ArXiv Preprint, 2021, ArXiv:2104.00298.

[21] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. ArXiv Preprint, 2020, ArXiv:2010.11929.

作者简介

赵亮,硕士生导师,主要研究方向为深度学习、智慧医疗。
E-mail:zhaoliang_270@163.com

付园坤,硕士研究生,主要研究方向为深度学习、图像处理、智慧医疗。
E-mail:yuankun_Fu@163.com

陈涵欣,本科生,主要研究方向为深度学习、图像处理、智慧医疗。
E-mail:1469537070@qq.com

魏政杰,硕士研究生,主要研究方向为深度学习、智慧医疗。
E-mail:wcizj0328@126.com

云晴,硕士研究生,主要研究方向为深度学习、图像处理、智慧医疗。
E-mail:2297337453@qq.com

金军委,讲师,主要研究方向为深度学习、智慧医疗。
E-mail:jinjunwci24@163.com